



Genetic footprinting of the HIV co-receptor CCR5: delineation of surface expression and viral entry determinants

Ricardo Quinonez,^a Indu Sinha,^a Ila R. Singh,^b and Richard E. Sutton^{a,*}

^a Department of Molecular Virology and Microbiology, Center for Cell and Gene Therapy, Baylor College of Medicine, Houston, TX 77030, USA

^b Department of Pathology, Columbia University College of Physicians & Surgeons, New York, NY 10032, USA

Received 28 June 2002; returned to author for revision 26 September 2002; accepted 2 October 2002

Abstract

Human immunodeficiency virus type 1 (HIV-1) utilizes CD4 as a primary receptor for viral entry and any of several 7-transmembrane chemokine receptors, including CCR5, as a co-receptor. Previous studies have demonstrated that multiple extracellular domains (ECDs) of CCR5 contribute to co-receptor function; here we applied genetic footprinting to CCR5 to confirm and extend those investigations. In genetic footprinting, a duplex oligonucleotide is inserted into the DNA sequence of interest by use of either a bacterial transposase or retroviral integrase. Here, CCR5 mutants were analyzed in bulk for their ability to be expressed on the recipient cell surface and to mediate viral entry of R5 HIV isolates. Most of the approximately 150 CCR5 mutants were not expressed on the cell surface. Of those remaining, 8 were specifically reduced or absent after macrophage (M)-tropic HIV infection, confirming a critical role of ECDs three (extracellular loop 2 or ECL2) and possibly four (ECL3) in viral entry. Mutational and functional analyses of ECD4 (ECL3) suggest it is under severe topological constraint for CCR5 surface expression and are consistent with it contributing to co-receptor function.

© 2003 Elsevier Science (USA). All rights reserved.

Introduction

Human immunodeficiency virus type 1 (HIV-1), along with the other primate lentiviruses, utilizes a primary receptor (CD4) and any of a growing number of co-receptors to mediate viral entry into susceptible cells (for reviews, see Berson and Doms, 1998; Littman, 1998; and Berger et al., 1999). The HIV-1 co-receptors belong to the family of chemokine receptors, 7-pass transmembrane G-protein coupled proteins whose natural functions include stem cell homing, lymphocyte trafficking, and organ development (for reviews, see Rossi and Zlotnik, 2000; and Sallusto et al., 2000). The major co-receptors are considered to be CXCR4 and CCR5, which function for T cell (T)-tropic (X4) and macrophage (M)-tropic (R5) HIV isolates, respectively.

Over the last several years, several mutagenesis studies of CCR5 have been performed, including functional analyses of human-mouse and chemokine receptor chimeric mol-

ecules (Atchison et al., 1996; Rucker et al., 1996; Bieniasz et al., 1997; Doranz et al., 1997; Picard et al., 1997), alanine-scanning mutagenesis (Dragic et al., 1998; Farzan et al., 1998; Rabut et al., 1998), and targeted mutagenesis (Doranz et al., 1997; Ross et al., 1998), although the latter studies have not been particularly informative. The results of these investigations have highlighted the importance of the first extracellular domain (ECD1) of CCR5 (also known as the amino terminal domain or Nt) for binding to gp120 of both R5 and dual-tropic (R5X4) viruses (reviewed in Berger et al., 1999; and Dragic, 2001). For example, chemokine receptor chimeras utilizing Nt from CCR5 are functional as R5 co-receptor. In addition, negatively charged (Asp-2, Asp-11, and Glu-18) and tyrosine residues (Tyr-3, -10, -14, and -15) are critical for both binding to gp120 and viral entry. Multiple other residues in ECD1 have also been demonstrated to contribute to co-receptor function, including Ser-6, -7, Ile-9, Asn-13, Gln-21, and Lys-22.

The other ECDs, especially the third (ECD3, also known as the second extracellular loop or ECL2), contribute to viral entry, in complex and envelope-specific manners. Gly-163 (thought to be located in the fourth transmembrane

* Corresponding author. Fax: +1-713-798-3586.

E-mail address: rsutton@bcm.tmc.edu (R.E. Sutton).

domain), Tyr-184, Ser-185, Arg-197, all have been shown to affect co-receptor function by mutagenesis studies. The importance of ECD3 for viral entry (but not for gp120 binding) has been buttressed by studies showing inhibition of virus-cell fusion using monoclonal antibodies that recognize epitopes in ECL2 (Wu et al., 1997; Lee et al., 1999; Olson et al., 1999). This suggests that ECD3 may play a role after gp120 binding to the Nt, perhaps during conformational changes of CCR5 or co-receptor oligomerization.

Of note, the fourth ECD (ECL3) is absolutely conserved between man and mouse and has not been subjected to as extensive mutational analysis. However, limited evidence suggests the importance of Asp-276 and Gln-280 (Doranz et al., 1997; Farzan et al., 1998). For the former, reduced co-receptor function was only observed in the context of other mutations (namely Asp-11 and/or Arg-197) and only for R5X4 and not R5 viral envelopes. In fact, in one study D276A failed to express at detectable levels as measured by flow cytometry, and cell surface expression of Q280A was consistently lower than that of wild-type CCR5 (Farzan et al., 1998). Other point mutations in ECD4 appear to have little effect on co-receptor function (Doranz et al., 1997; Farzan et al., 1998). We sought to confirm and extend the results of the mutagenesis studies by the use of genetic footprinting.

Genetic footprinting is a saturation mutagenesis technique in which either a bacterial transposase or retroviral integrase is used in vitro to randomly insert a duplex oligonucleotide by a concerted integration event into a DNA sequence of interest (Singh et al., 1997). The entire collection or library of mutant sequences is then subjected to a functional selection and analyzed by PCR using a specific property of the inserted oligonucleotide. This method thus allows parallel analysis of hundreds if not thousands of mutants without isolating a single one individually. Genetic footprinting has been applied with success to the bacterial suppressor tRNA SupF (Singh et al., 1997), cis acting nucleic acid sequences of the HIV-1 genome (Laurent et al., 2000), and the Maloney murine leukemia virus (MLV) envelope glycoprotein (Rothenberg et al., 2001). It has yet to be used to study the coding sequences of a cellular eukaryotic gene.

Here we report the application of genetic footprinting to CCR5. A library of oligonucleotide insertions was made using MuA transposase in a hemagglutinin (HA) epitope-tagged version of CCR5 already present in a MLV-based vector. The library was used to transduce cells bearing CD4 and several functional selections were performed. A majority of the mutant CCR5 proteins were not expressed on the cell surface, suggesting that membrane trafficking of CCR5 is quite sensitive to its structure. Of the mutants that were expressed on the cell surface, 8 were still present after transduction of the cells with HIV pseudotyped with vesicular stomatitis virus (VSV) G protein but greatly reduced in amount after transduction of cells with HIV pseudotyped with the envelopes of the R5 isolates ADA, BaL, and

SF162. The location of these insertions confirms the critical role of the third and possibly fourth ECDs (ECL2 and ECL3, respectively) for CCR5 co-receptor function, which was corroborated by specific mutants of the fourth ECD. These results thus extend the general utility of genetic footprinting to the functional analysis of protein-coding eukaryotic genes.

Results

Genetic footprinting strategy

Fig. 1 outlines the strategy for genetic footprinting of CCR5. The gene of interest (CCR5) in an MLV-based plasmid vector pBabeCCR5-HA (Liu et al., 1996) was subjected to a concerted integration reaction in vitro using purified MuA transposase (Fig. 1A, B). After oligonucleotide addition, the resulting DNA was cut with Mlu I, religated, and transformed into *E. coli*. Approximately 4.5×10^5 bacterial colonies were pooled and plasmid DNA prepared. Randomness of integration was confirmed by digesting the DNA with both Not I (cleaves the vector backbone once) and Mlu I, which resulted in a smear of DNA fragments as judged by horizontal agarose gel electrophoresis. Greater than 95% of the plasmids had a single Mlu I site. Note that the coding sequence of CCR5 is only approximately 1.1 kb in length, so roughly 17% of the oligonucleotide inserts were inserted within CCR5 (the rest within vector sequences).

Purified plasmid DNA along with pHIT60 (encodes MLV Gag-Pol) (Soneoka et al., 1995), and pMEVSV G (encodes VSV G) (Sutton et al., 1998) was used to transiently transfect 293T cells (Fig. 1B). The resulting pseudotyped viral supernatants were titered on HOS.T4 cells by end-point dilution, and the titers were approximately 6×10^4 IU/ml. Forty milliliters of viral supernatant (2.4×10^6 IU) were used to transduce 5×10^7 HOS.T4 cells so that the MOI was $\sim .05$ and double transductants would be avoided. This low MOI was important to reduce any type of trans-complementation in the targets. At the same time, since this amount of IU represents only ~ 1000 MLV vector clones (see below), each is represented on average 2400-fold. Even if some are under-represented 50-fold, those clones would still be represented by ~ 50 different transduction events, thus greatly reducing or eliminating any clonal bias in transgene expression which is well-known to occur after retroviral transduction. After puromycin selection, cells were pooled, washed, expanded for several weeks, and frozen in aliquots of at least 10^7 cells to avoid bottle-necking and loss of representation. These cells were then used for functional tests.

Fig. 1C demonstrates the principle behind genetic footprinting. The gene of interest has in each case no more than a single duplex oligonucleotide inserted at a random position. Using two PCR primers (one labeled with P-32 and the other one

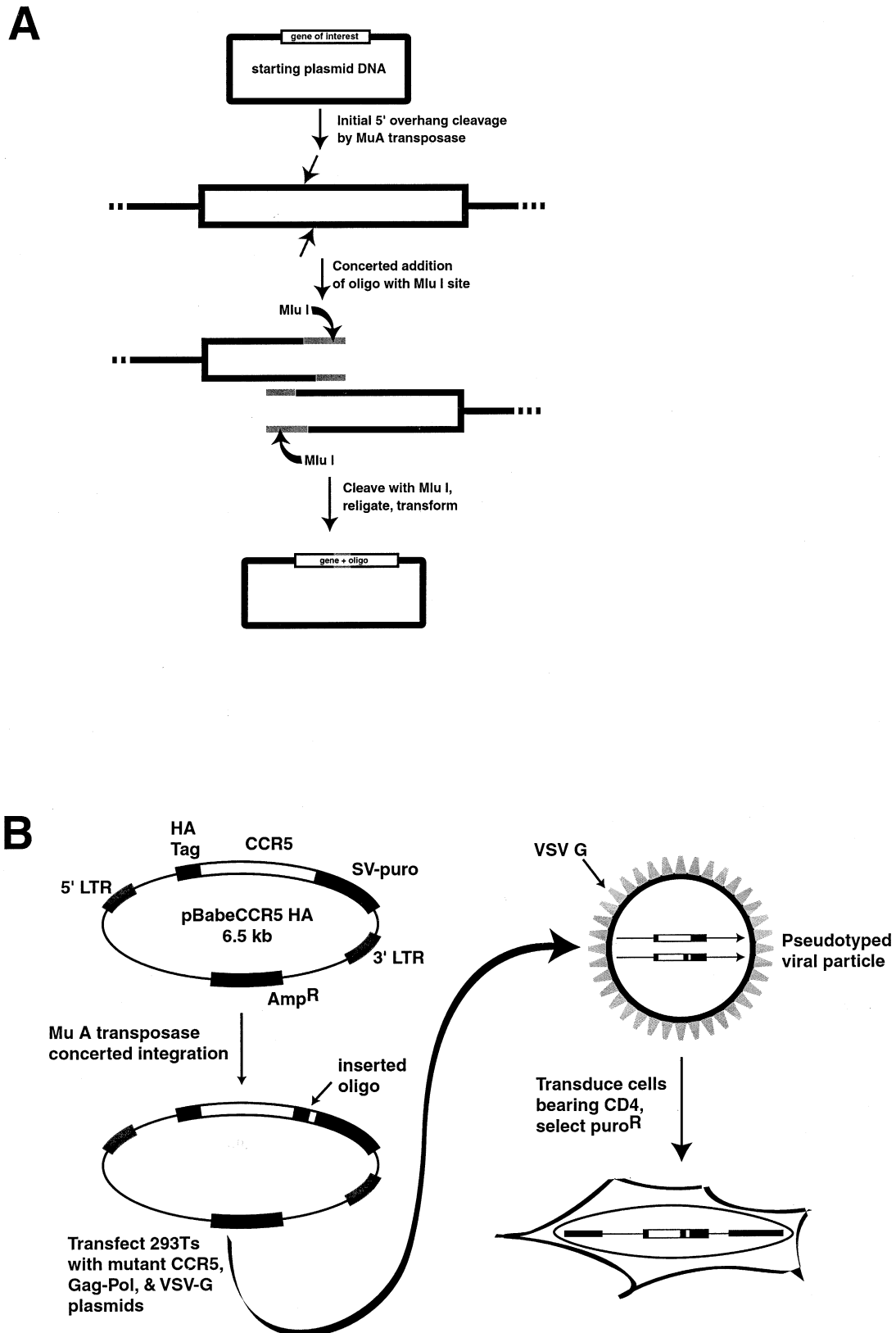


Fig. 1. Schematic of genetic footprinting. (A) shows MuA transposase-mediated concerted integration of the duplex oligonucleotide containing the Mlu I site into the gene of interest contained on plasmid DNA. Note, however, that the integration event may occur anywhere in the plasmid DNA. (B) shows the MLV-based vector encoding HA-tagged CCR5. This was subjected to concerted integration and the plasmid library was co-transfected into 293T cells to produce VSV G-pseudotyped MLV particles (shown with a diploid RNA genome), which were then used to transduce HOS T4 cells. The puro^R cells were then sorted and subjected to functional selections. (C) shows the basic principle of the PCR technique to determine the point of oligonucleotide insertion. The oligo is represented by a closed box (region where function is not disrupted) or open box (function is disrupted). B = biotin, UN = unselected population, SEL = selected population. (D) shows some of the functional selections performed here. At top are HOS.T4 cells, each with a single MLV integrant. Within the nucleus is the HA tagged CCR5 gene, with the inserted oligonucleotide as a black box. Note after sorting for surface expression, CCR5 (grey circles) is on the cell surface. After infection with HIV- blasti (ADA), only one of the cell clones remains. At bottom is a schematic of the denaturing polyacrylamide gel, showing idealized results, with square brackets indicating regions of functional importance.

frame 0

STOP ↓

N₁N₂N₃N₄N₅TGACGCGTCAN₁N₂N₃N₄N₅

↑

Met
Leu ↓
Val Thr Arg His
Glu

frame 1

N₁N₂N₃N₄N₅TGACGCGTCAN₁N₂N₃N₄N₅

↑

not: Met
Glu
Lys
Gln ↓
Trp Asp Ala Ser

frame 2

N₁N₂N₃N₄N₅TGACGCGTCAN₁N₂N₃N₄N₅

↑

Diagram illustrating the effect of a 32 bp deletion in the CCR5 coding sequence. The top line shows the wild-type sequence with ATG and TGA start/stop codons. The middle line shows the deletion, and the bottom line shows the resulting reading frame shift, which leads to a premature stop codon (TGA) at the end of the sequence.

biotin), the appropriate PCR product is amplified, isolated, and in this case subjected to Mlu I restriction endonuclease cleavage. Restriction digest products are separated on a denaturing polyacrylamide sequencing gel and exposed by autoradiography. All insertions will be revealed in the unselected material. Once a functional selection has been applied, some products will no longer be present, thus delineating the region of the gene required for function (bracket in Fig. 1C).

After blasticidin selection, resultant colonies were pooled, expanded, and genomic DNA prepared. Anticipated results are shown at the bottom of Fig. 1D. As expected, there were no changes in the pattern of the radioactive PCR products when comparing the initial plasmid DNA pool to that of the post-MLV transduction of the HOS.T4 cells (data not shown). However, there was marked loss of products after sorting for surface expression of CCR5 (see below). This loss would thus define regions important for cell surface expression of CCR5. Unexpectedly, there was some additional loss of bands after HIV-blasti (VSV G) transduction (see below). However, there was additional product loss after transduction of the R5 envelope-bearing viruses (see below). This loss would define regions or amino acid residues critical for infection with these R5 viruses.

Fig. 2A illustrates the different reading frames of the inserted oligonucleotide. The Mlu I site is centrally located within the new 10 bp insertion and the 5 bp duplication. This

5–5 bp duplication occurs greater than 90% of the time for the concerted MuA transposase *in vitro* integration reaction (R. Crowley and P.O. Brown, unpublished results). Thus a total of 15 bp are inserted so that the reading frame is maintained. Unfortunately, unavoidable DNA recognition sequence constraints of the MuA transposase imposed a stop codon (TGA) if the oligonucleotide was inserted in frame. This would then lead to protein truncation at the point of insertion. Note that for the other two reading frames the predicted codons would lead to incorporation of a mix of hydrophobic, basic and acidic amino acid residues. In both cases there would also be duplication of one residue. It is also important to realize that as shown in Fig. 2A the insertion point can either be considered to be at position N5 or N1. Both interpretations are technically correct, but since proteins are translated starting at the amino terminus we have regarded N5 to be the insertion point and the second N1–N5 the duplicated sequence.

When determining the point of oligonucleotide insertion, the same radioactive DNA primer was used in both the PCR and the Sanger sequencing reaction and the homologous DNA sequence was electrophoresed on neighboring gel lanes. Note that when the upper (5') PCR primer was radioactive, after Mlu I digestion the radioactive product was 3 bp longer than the point of insertion, and in the case of the lower (3') PCR primer, the radioactive product was 8 bp longer than the true point of insertion (N5 in Figure 2A). From the original plasmid library we isolated approximately 10 mutants and sequenced the site of oligonucleotide insertion. In all of the examined mutants, the expected 5 bp duplication was present at the point of insertion. Although none of the mutants were informative (*i.e.*, they were lost after bulk cell-surface sorting), the point of insertion agreed with that determined by the PCR and denaturing sequencing gel method.

Fig. 2B illustrates the PCR strategy. Using the genomic DNA, an initial product of 1.4 kbp was amplified using the external primers. This 1.4 kbp product was used in nested PCR reactions, with the approximate positions of the primer pairs and product sizes as shown. There is overlap between all PCR products (although somewhat minimal between products 4 and 5) and no product is longer than approximately 450 bp. Thus, the point of insertion could be identified in each PCR product by switching the position of the radioactive and biotinylated primers and performing two or three loads on a 40 cm denaturing polyacrylamide sequencing gel. In the regions of the PCR product where the point of insertion could be read using either radioactive primer, the two results agreed perfectly, confirming the 5–5 duplication rule as described above.

Multiple regions of CCR5 contribute to surface membrane trafficking

After transduction of the mutant CCR5 library into HOS.T4 cells, cells were magnetic bead-sorted for surface expression of CCR5. Sorted cells were expanded and transduced with the different vector supernatants as described

above. To confirm that the mutant CCR5 proteins were actually being expressed on the cell surface, we performed flow cytometry on the different populations. As shown in Figure 3A, 30% of the unsorted cells were judged to be positive, whereas after sorting and the different transductions between 52% and 84% were positive. We are at a loss to explain why these numbers are low (*i.e.*, not 100% after R5 viral infection), but it appears to be due to insufficient increase in mean cell fluorescence after antibody staining (see Figs. 3E and 3F).

Because we had noted extensive band loss or “dropout” after magnetic bead sorting in preliminary experiments, we became concerned about loss of representation or “bottlenecking” despite the fact that 4×10^7 cells had been sorted. We thus sorted three batches of cells in parallel and performed genetic footprinting on the genomic DNA. Fig. 4 shows that essentially the same band dropout was observed with all three sorted samples (compare unsorted vs. sorted-1,2,3). This result was confirmed with two other primer pairs (data not shown). Thus, the differences that we had initially observed between the sorted and unsorted populations were unlikely to be due to misrepresentation but instead to true dropout, consistent with loss of function.

This extensive band loss was observed with other primer pairs throughout CCR5 (Fig. 5 and 6 and data not shown), suggesting that cell surface expression of CCR5 was quite sensitive and easily disrupted by the 15 bp oligonucleotide/5 amino acid residue insertion. As summarized in Fig. 7, out of a total of 148 oligonucleotide insertions (excluding known insertions within the HA epitope), only 53 (36%) were consistently observed in the sorted population. Note many oligonucleotides (37) were inserted in frame to create a truncated gene product and were only present in the unsorted population (grey residues in Fig. 8). These were present in ECDs 1, 3, and 4, all transmembrane regions, and most of the intracellular domains.

There were also many oligonucleotide insertions (58) that were out of frame but still disrupted surface expression of CCR5 and thus were not present within the sorted population (see Fig. 7 and red residues in Fig. 8). These insertions were present in all extracellular, transmembrane, and intracellular domains of CCR5. This suggests that the 7-transmembrane topology or cell surface trafficking of CCR5 is quite sensitive to a small 5 amino acid residue insertion, even if it is completely within an intracellular or extracellular domain.

A number of oligonucleotide insertions (15, equivalent to 10% of the mutants) were present in the sorted and unsorted populations but not in the post-HIV(VSV G) transduction population (see Figures 4–6 and green residues in Fig. 8). It is possible that this reflects mutant loss prior to HIV(VSV G) infection since these insertions were also absent in all three of the post-R5 transduction populations. The great majority of these insertions (13/15) were in the transmembrane regions. Whether these mutants somehow represent PCR artifacts or perhaps disruption/modification of the cell

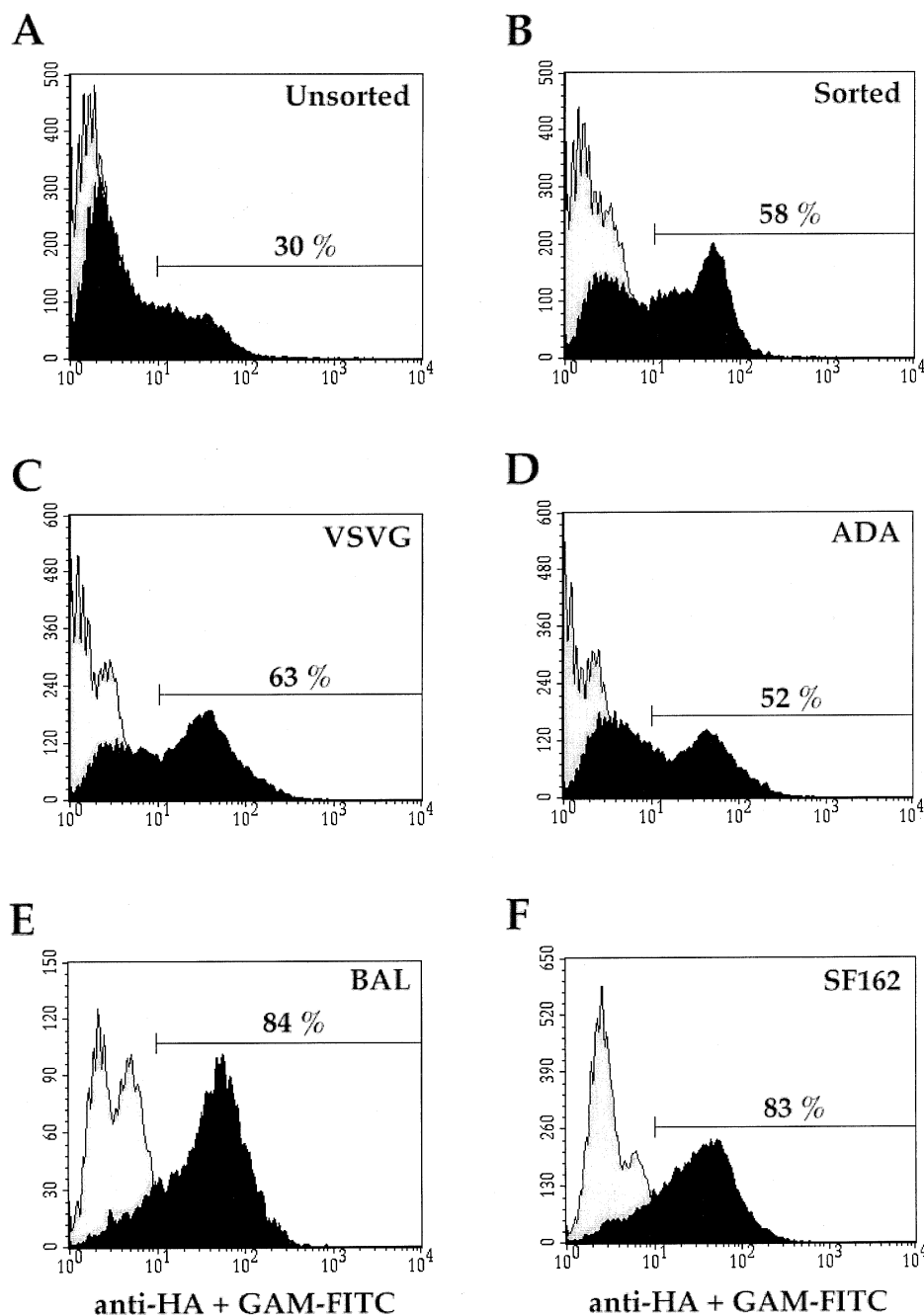


Fig. 3. Cell surface expression of CCR5. The six different HOS.T4 cell populations were analyzed by flow cytometry using an indirect method for surface expression of CCR5. (A) Post-MLV transduction and unsorted, (B) post-magnetic bead sorted, (C) post-HIV (VSV G) transduction, (D) post-HIV (ADA) transduction, (E) post-HIV (BaL) transduction, and (F) post-HIV (SF162) transduction. Grey histograms: goat anti-mouse-FITC (GAM-FITC) alone added, black histograms, 12CA5 and GAM-FITC added. Bars and percentages indicate positive population (gated at ~1% background).

surface membrane such that not even VSV G pseudotypes are infectious is unknown; since these mutants were not present in any of the post-infection populations they did not contribute to or interfere with the analysis of CCR5 function.

CCR5 is tolerant to oligonucleotide insertions

It was already known that short peptide epitopes could be placed within the first ECD of CCR5 and still retain co-

receptor function. For example, the starting CCR5 construct for these investigations has an eight amino acid HA epitope tag inserted at residue 14 (Liu et al., 1996). However, even placement of this epitope proved difficult since the importance of the tyrosine residues in that first domain were unappreciated at the time of vector construction (N. Landau, personal communication). This genetic footprinting procedure also identifies points of oligonucleotide insertion that

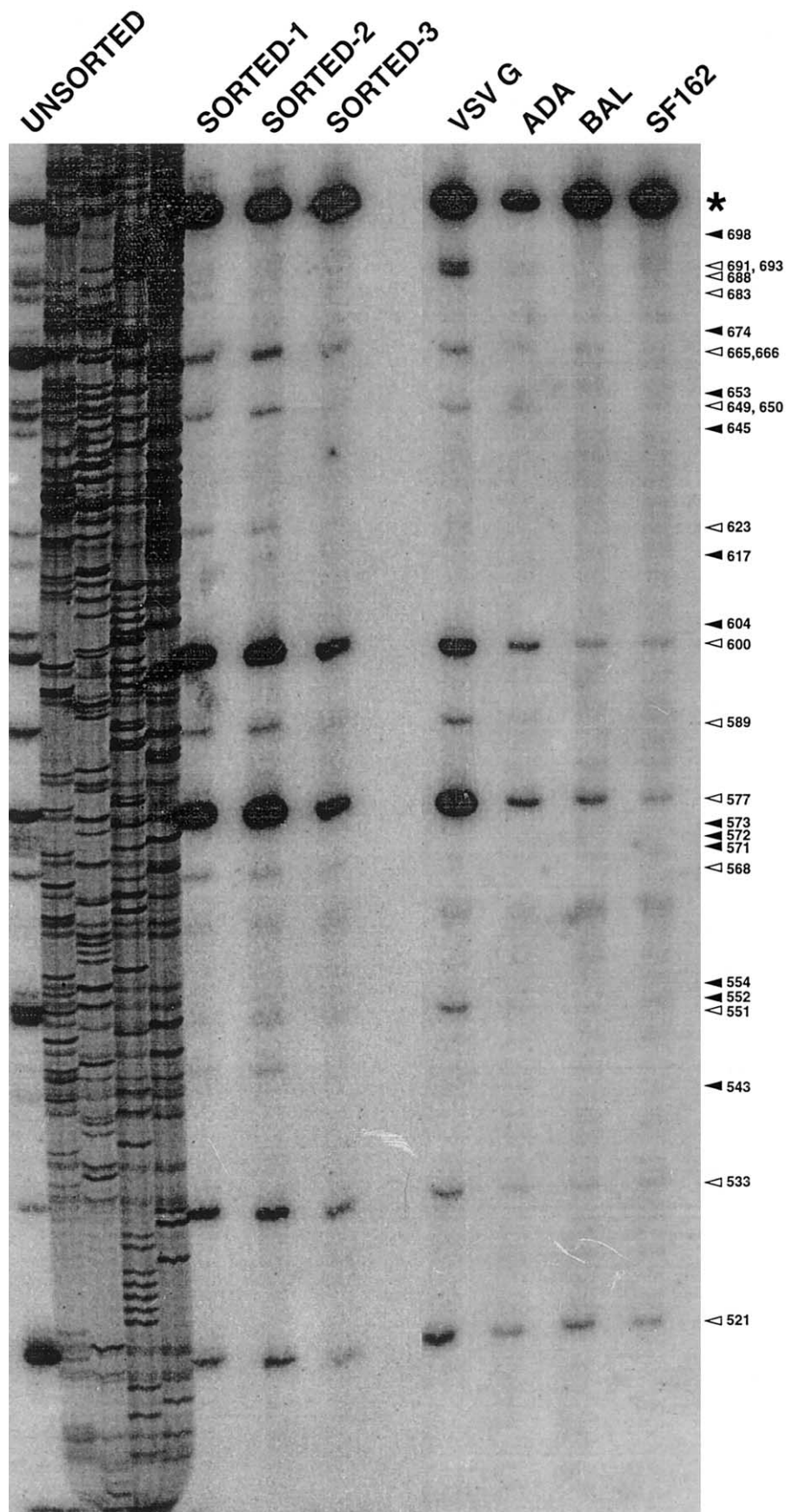


Fig. 4. Genetic footprinting of CCR5, region 406–716. Genomic DNA was prepared from the cell populations shown in Figure 3 and nested PCR performed. In this case, the 627U primer was labelled with P-32. After Mlu I digestion, products were size fractionated on a denaturing polyacrylamide gel along with the homologous sequence and subjected to autoradiography. Sorted-1, Sorted-2, Sorted-3 are the three cell samples sorted in parallel for surface CCR5 expression. * is the full-length, undigested PCR product (311 nt), closed arrowheads indicate products observed in the unsorted population only, and open arrowheads indicate products observed in the sorted population as well. Refer to Figure 7 for bp numbering.

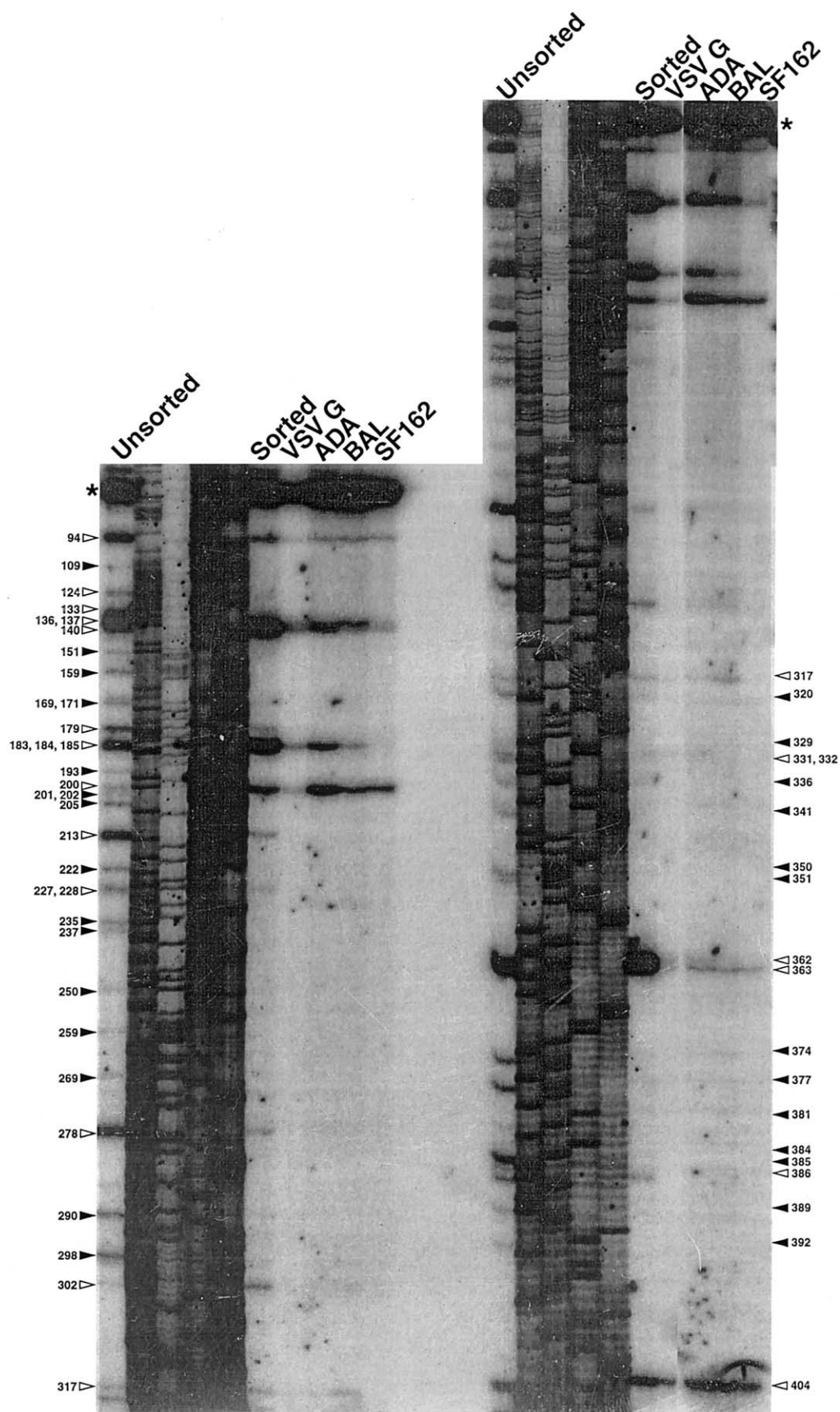


Fig. 5. Genetic footprinting of CCR5, region 60–496. Here, the 693B primer was labeled with P-32. Please refer to Figure 4 for experimental details and legend key.

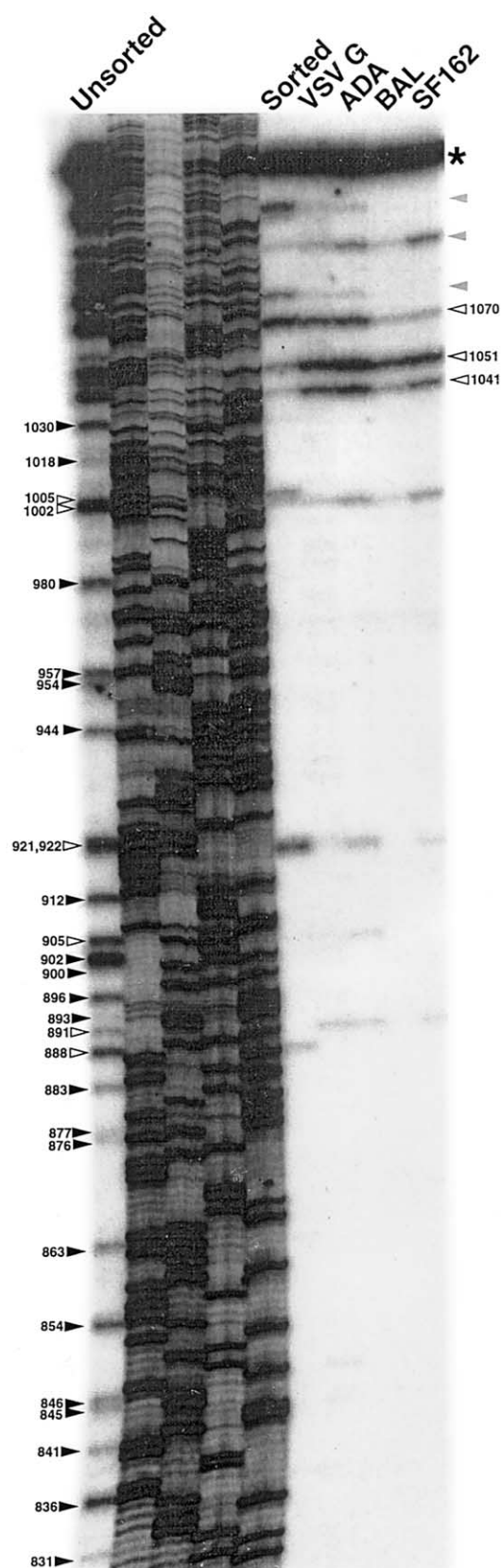


Fig. 6. Genetic footprinting of CCR5, region 686 to C-terminus. In this case, the 907U primer was labelled with P-32. Please refer to Figure 4 for experimental details and legend key. Grey arrowheads indicate insertions 3' of the stop codon at position 1075.

are tolerated by CCR5 (depending upon the required functional tests, of course).

Multiple oligonucleotide insertions (26) were present in all the cell populations analyzed: unsorted, sorted, post-HIV(VSV G), and post-R5 viral transductions (see Figs. 4–6 and blue residues in Fig. 8). Surprisingly, nearly half of these were in the transmembrane regions. There were 0–2 insertions in most of the extracellular and intracellular domains, with the exception of the cytoplasmic tail, where there were 6 tolerated oligonucleotide insertions. This is not unexpected, since the cytoplasmic tail is not required for co-receptor function (Alkhatib et al., 1997) and C-terminal green fluorescent proteins are fully active (D. Littman, personal communication). Whether all of these would be still be present if other functional tests were to be applied (e.g., calcium/cellular signalling, chemokine binding, or cellular chemotaxis) is unknown.

Residues in the third and fourth ECD are critical for CCR5 co-receptor function

Of the 148 mutant CCR5 proteins, only 8 were present in the unsorted, sorted, and post-HIV(VSV G) genomic DNA material, but reduced or absent in the post-R5 envelope genomic DNA. These oligonucleotide insertions were found within the third ECD (ECL2), the fourth transmembrane region, and the third intracellular loop (Figure 4 and yellow residues in Figure 8). The first Cys in the third ECD (residue 184 here) is thought to make a critical disulfide bond linkage with the Cys in the second ECD (residue 107 here, see Figure 8). This bond is likely important for the proper folding and function of CCR5. Note one oligonucleotide insertion at C184 led to protein truncation and was only present in the unsorted population, whereas the other insertion was at bp 551 and was specifically absent from the post-R5 envelope transduction populations (Figure 4). This insertion results in the amino acid sequence . . . Tyr-Thr-Cys-Asp-Ala-Ser-Thr-Cys-Ser-Ser . . . , where the underlined portion are the novel residues. Although it appears to be expressed on the cell surface, this protein may be misfolded because of the presence of a second Cys or alternatively critical determinants for CCR5 co-receptor function may be disrupted by this insertion.

The second oligonucleotide insertion specifically reduced in the post-R5 envelope transduction populations was within the Tyr residue 193. This insertion results in the amino acid sequence. . . Ser-Gln-Leu-Thr-Arg-His-Gln-Tyr-Gln . . . , where the novel residues are underlined. Although expressed on the cell surface, this mutant CCR5 was reduced by >90% after HIV(ADA), HIV(BaL), and HIV(SF162) transductions. Presumably it too disrupts critical co-receptor determinants.

The third and fourth oligonucleotide insertions specifically reduced in the post-R5 envelope transduction populations led to truncated proteins after the Trp residue 196 and the Gln residue 200, respectively. Since both of these pro-

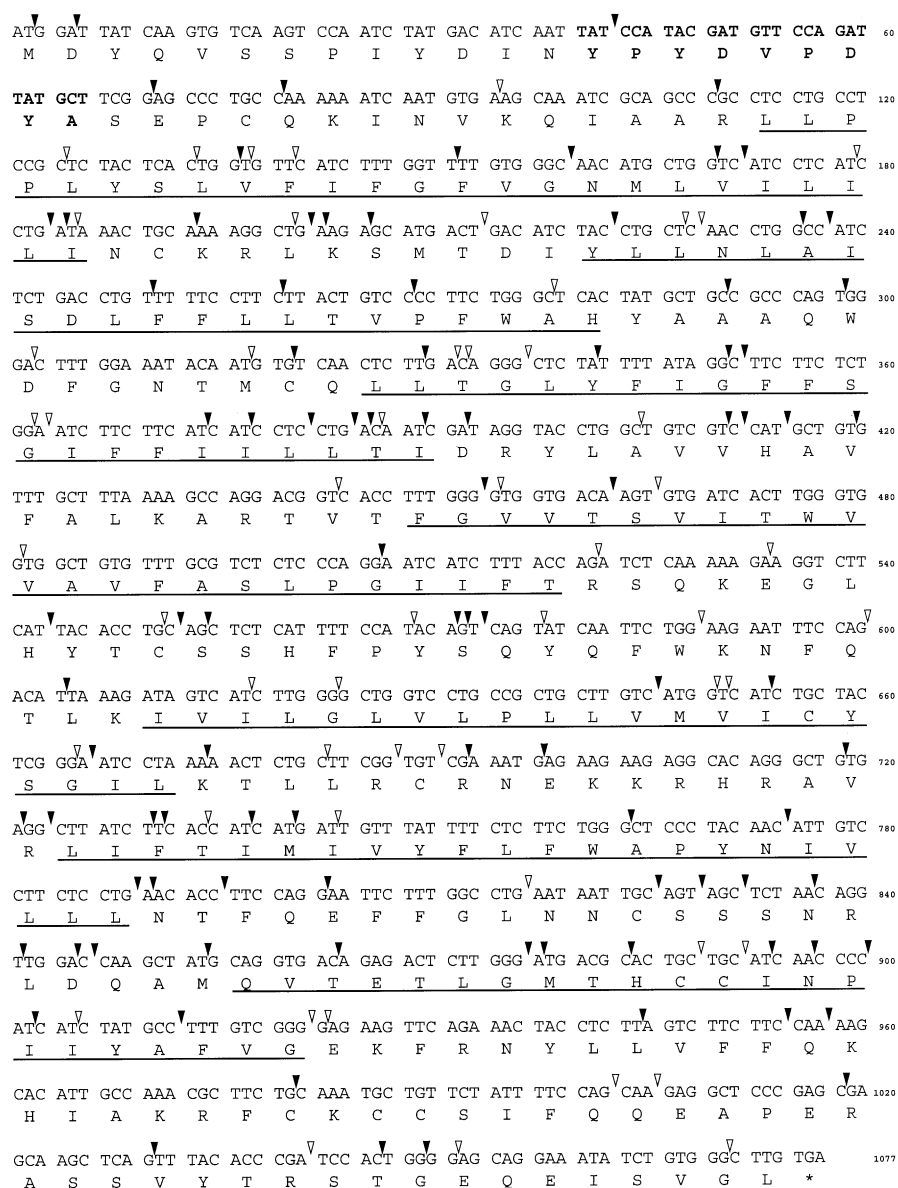


Fig. 7. Points of insertion of the oligonucleotide. Closed arrowheads indicate products only observed in the unsorted population, and open arrowheads indicate products observed in the sorted population as well. Bold type indicates the HA tag. Note the first residue of the epitope (Y) is also part of CCR5. In addition, two residues (Y15 and T16) of wild-type CCR5 were removed during the addition of the epitope, which did not appear to compromise co-receptor function. Underlined residues indicate putative transmembrane domains.

teins were expressed on the cell surface, they likely have altered conformations that will not support co-receptor function. Alternatively, there may be critical determinants of co-receptor function that are C-terminal to these residues (see below).

The fifth and sixth oligonucleotide insertions variably present within the post-R5 envelope transduction populations were at nucleotide positions 649 and 650 within the Val residue 217 (thought to lie within the fifth transmembrane domain) (Figs. 4 and 8). These insertions gave rise to inserted peptides. . . Val-Met-Val-Thr-Arg-His-Met-Val-Ile. . . and. . . Val-Met-Val-Asp-Ala-Ser-Met-Val-Ile. . . , respectively, where the underlined portion is the

novel sequence. It is possible that the presence of the single charged residue (Arg or Asp) within the transmembrane domain disrupted conformation to such an extent these were reduced in the R5 envelope populations (despite the fact they are trafficked to the cell surface).

The seventh and eighth oligonucleotide insertions absent within the post-R5 envelope transduction populations were at positions 688 and 691 (Fig. 4). These insertions led to protein truncations after amino acid residues Arg 229 and Cys 230, respectively (third intracellular domain, Fig. 8). These insertions were somewhat unusual in that they were present in the unsorted population, consistently reduced in the post-sorted population, quite prominent in the post-

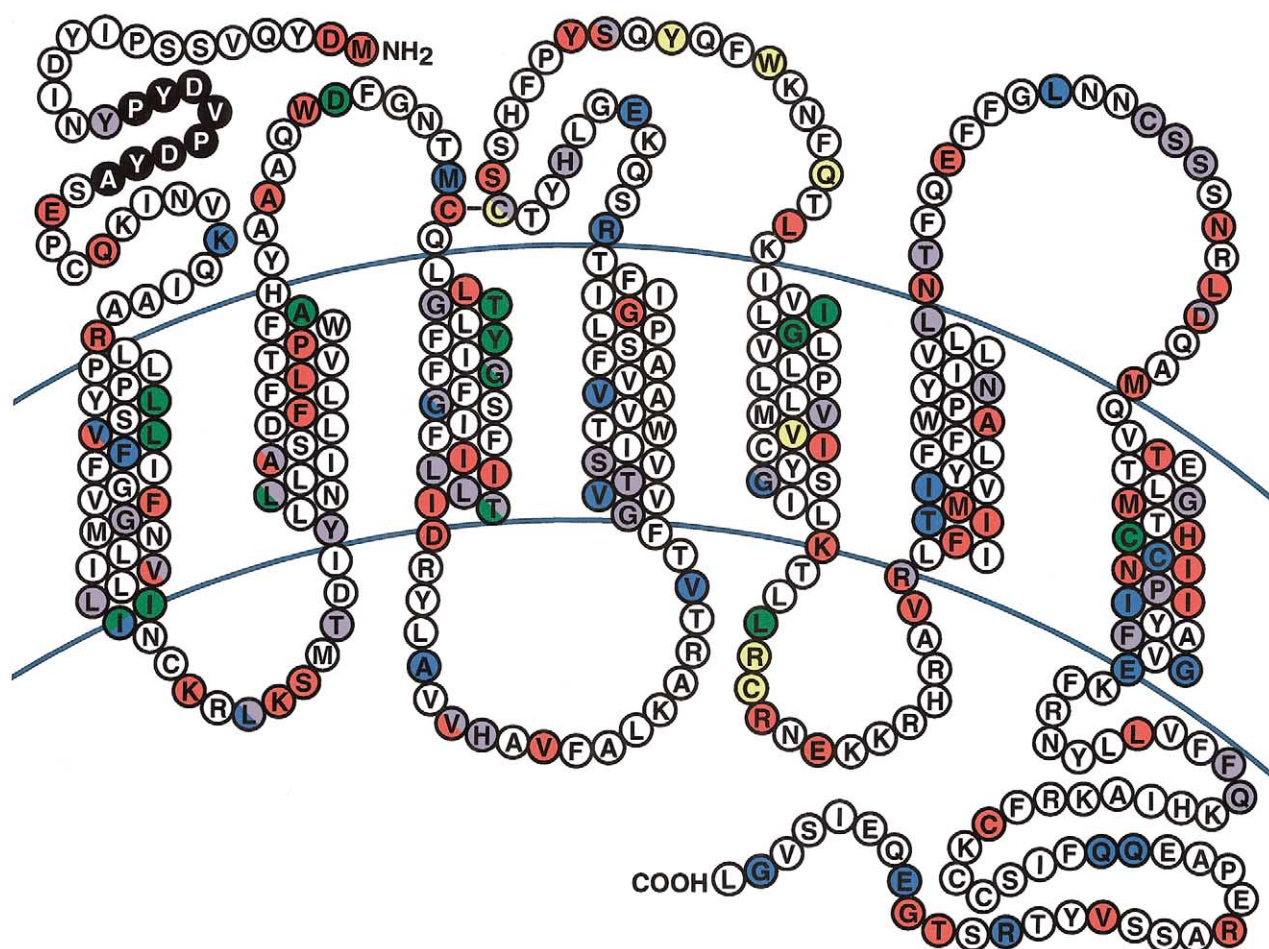


Fig. 8. Schematic of CCR5 structure. Plasma membrane is indicated by the curved cyan lines; ECDs are at top, intracellular loops at bottom (amino and carboxy termini indicated). Black residues indicate the HA epitope tag, grey residues truncations at that position present only in the unsorted population, red residues non-truncating insertions at that position present in the unsorted population only, blue residues insertions at that position present in all populations, green residues insertions at that position present in the unsorted and sorted populations only, and yellow residues insertions at that position present in the unsorted, sorted, and post- HIV(VSV G) but absent in the post-R5 envelope populations. Split circles indicate multiple insertions at that residue.

HIV(VSV G) transduction population, and virtually absent in all three of the post-R5 transduction populations (Fig. 4). It is unknown whether C230 forms a disulfide bond with any of the other intracellular Cys residues. These two oligonucleotide insertions both suggest that there are critical CCR5 co-receptor determinants C-terminal to C230, quite possibly within the fourth ECD.

To address this question more directly, we made several mutations in the highly conserved fourth ECD (Table 1). Some of these added one to a few residues, others deleted a few to most of the ECD. Of note, only one of the mutants was expressed on the cell surface (4ECD-6), although the percentage of positive cells, as judged by flow cytometry, was reduced 20-fold compared to wt control (Fig. 9). The positive mutant CCR5 cells, however, had the same, if not greater, mean fluorescence intensity (MFI) as wt control, whereas the titer of HIV-Neo (ADA) on 4ECD-6 expressing cells was reduced more than 5,000-fold compared to wt control (Table 1).

Discussion

Utility of genetic footprinting

We have described here the use of genetic footprinting to analyze the function of a cellular eukaryotic gene. Previously, the use of this technique has been confined to the study of a bacterial suppressor tRNA, the cis-acting sequences of the HIV genome, and a retroviral envelope. The study of CCR5 co-receptor function adds a level of complexity since specific amino acid residues were introduced and multiple complex functional assays performed. Because we were analyzing insertions in mammalian genomic DNA, the use of PCR was unavoidable and we cannot exclude representational bias after nested amplification. However, when comparing nested PCR performed on the original plasmid DNA pool vs. the genomic DNA of the unsorted population, there was no clear representational loss or gain, suggesting that the points of oligonucleotide insertion and their relative intensities after PCR reflect

Table 1
CCR5 fourth extra-cellular domain mutants

Mutant ^a	Expression ^b	Titer ^c
wt ECD4		
aac acc ttc cag gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct atg N T F Q E F F G L N N C S S S N R L D Q A M	+++	~5000
4ECD-1		
aac acc ttc cag gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct ggg cat atg N T F Q E F F G L N N C S S S N R L D Q A G H M	–	ND
4ECD-2		
aac acc ttc cag gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct cat atg N T F Q E F F G L N N C S S S N R L D Q A H M	–	ND
4ECD-3		
aac acc ttc cag gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct cat atg N T F Q E F F G L N N C S S S N R L D Q A H M	–	ND
4ECD-4		
aac acc ttc cag gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct atg N T F Q E F F G L N N C S S S N R L D Q A M	–	ND
4ECD-5		
aac acc ttc cag gaa ttg tgc act gaa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct atg N T F Q E L S T E F F G L N N C S S S N R L D Q A M	–	ND
4ECD-6		
aac acc ttc cag gaa ttc agt cga caa ttc ttt ggc ctg aat aat tgc agt agc tct aac agg ttg gac caa gct atg N T F Q E F S R Q F F G L N N C S S S N R L D Q A M	+	< 1

^a Bold residues indicate those added, boxed residues those deleted (compared to wt) from the fourth ECD, all sequence-confirmed.

^b Relative cell surface amounts, as judged by flow cytometry using an anti-CCR5-PE monoclonal antibody. +++ denotes 20–40%, +1–3%, and – < 0.1% positive cells.

^c IU/ml, using HIV-Neo (ADA) as determined on HOS cells also expressing CD4; ND not determined. For both wt and 4ECD-6, the HIV-IRES-eYFP (VSV G) titers were the same (~10⁷ IU/ml).

their true frequencies in the cell population and the corresponding genomic DNA.

After the initial amplification step, each region of the gene is subjected to PCR by a single set of primers under the same conditions and later digested with Mlu I; it is thus hard to ascribe the varying band intensities to PCR artifact. Increased or decreased intensity of any single cleaved PCR product most likely represents over and under-representation of that mutant in the library (as has been observed in other MuA transposase-generated libraries; see Laurent et al., 2000), respectively, although we have not formally shown that since our library sampling of specific mutants here was only ~6%.

The original mutant plasmid library had 450,000 members and was not liquid or plate amplified. Due care was exercised to avoid loss of representation or “bottlenecking”

by working with a large number of cells and pseudotyped particles at each step along the way (including the amount of genomic DNA for the nested PCR). It appears however, that the true complexity of the library was only ~1000 members. Thus, instead of ~70 oligonucleotide insertions per bp of the plasmid pBabeCCR5HA, on average there was 1 oligonucleotide insertion per 6 bp. Precisely why the library was 500-fold less complex than expected is a matter of speculation. It is unlikely to be due to loss of representation secondary to the multiple manipulations performed since the original mutant plasmid pool also demonstrated this frequency of oligonucleotide insertion. Rather, it may be due to the known marked site preference of MuA transposase, the bacterial integrase used here, which is reflected in the varying band intensities of the original library.

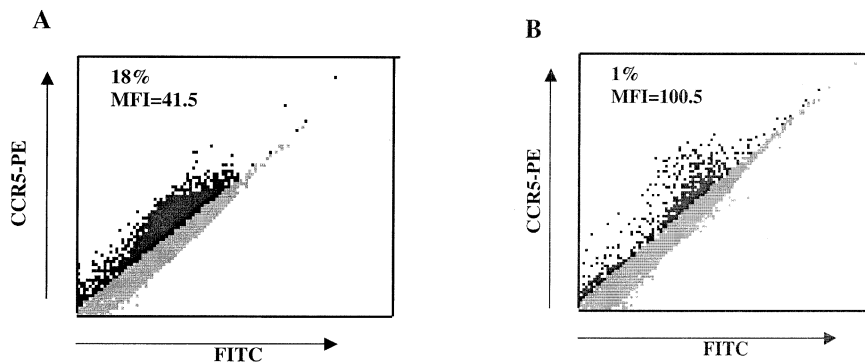


Fig. 9. Flow cytometry of HOS.CD4.CCR5 cells. HOS.CD4 cells were transduced with an MLV vector encoding either wt (A) or mutant ECD4-6 (B) CCR5. After puromycin selection, cells were stained for CCR5 using PE conjugated antibody. Gate was set so that < 0.1% of HOS.CD4 cells were positive (not shown). Percentage of gated cells and MFI (calculated as distance from origin) for each are indicated.

This insertional bias was previously observed when analyzing the cis-acting elements of the HIV genome, where the original library had similar complexity to the one described here and yet the frequency of oligonucleotide insertion was approximately 1 every 7 bp (Laurent et al., 2000). Similar bias was observed in the study of ecotropic MLV envelope (Rothenberg et al., 2001). The retroviral integrases do not appear to have such marked site preferences when naked DNA is used as a template (Brown et al., 1987; Pryciak and Varmus, 1992; Singh et al., 1997). Unfortunately, they are also more difficult to purify in soluble active form and have less intrinsic concerted integration enzymatic activity in comparison to MuA transposase.

We also endeavored to simplify genetic footprinting by not attempting to subclone the integration product but instead left it in the MLV-based vector. Previously, the integration products were cloned into a small plasmid vector or into the HIV genome (present within a plasmid vector). Thus, each member of the mutant plasmid library would have at most a single oligonucleotide insertion only within the region of interest but construction of the library is more laborious. Here, 15–20% of the members of the mutant plasmid library had a single oligonucleotide insertion within the coding sequence of CCR5 (the rest had none), but the library was relatively easy to construct since the linear concerted integration product was simply recircularized. Thus, to attain the same complexity as the subcloned library, the direct library would have to be 5-fold greater in size in this specific case. But since the true complexity of these libraries was only a few thousand (see above), this is readily accomplished. Achievement of equivalent complexity will obviously depend upon the size of both the plasmid and the region of interest. Direct libraries might be favored when the region of interest is large and the plasmid is small, whereas subcloned libraries are preferred when the region of interest is small and the plasmid large.

As in other studies using genetic footprinting, we did not attempt to reconstruct any of the specific CCR5 mutants in the library. Even with a complexity of only 1000 members, direct isolation of the mutants of greatest interest would be non-trivial, and as mentioned above the specific mutants isolated here were unfortunately all non-informative. Reconstruction of select mutants would also be difficult (although not impossible). Indeed, the strength of this method lies in the fact that the mutants are not analyzed individually but in bulk. When working well, the technique of genetic footprinting should allow the parallel analysis of thousand of mutants in any of a number of functional selections, limited only by the parameters of the biological system (see below).

Genetic footprinting is non-trivial. Complexity of the analyses grows substantially as the region of interest becomes larger and as the number of functional analyses increases. Cleaner results are obtained when the functional

selection is strong. Sorting for surface expression as described here is a moderately strong functional selection, although there is an unavoidable background (see Figure 3). Survival after lentiviral transduction is a strong functional selection since the untransduced cells die and the transduced cells are expanded logarithmically (and the background is almost non-existent). We did not attempt other functional selections of CCR5 such as G-coupled calcium signaling, chemokine ligand binding, or cellular chemotaxis since those selections are likely to be considerably weaker and not absolute.

Cell surface expression of CCR5

It is perhaps not surprising that approximately two-thirds (64%) of the oligonucleotide insertions resulted in a protein that did not get trafficked to the cell membrane, especially since many led to premature truncation. The well-known $\Delta 32$ allele causes a frame-shift mutation and a premature stop codon within the third ECD, resulting in a protein that is expressed in the cell but not on the cell membrane (Liu et al., 1996; Samson, 1996). What is more surprising is that half of the insertions reported here were in either the putative extracellular or intracellular domains, and half within the transmembrane regions (approximating the distribution of residues). This counter-intuitive result suggests that the topological constraints of the 7-transmembrane proteins might be much greater than previously thought.

Of course, insertion of five amino acids any where in CCR5 might greatly disturb trafficking or function (reflected here in the few mutants that survived the cell sorting). Efforts to develop “replacement” instead of “insertion” genetic footprinting libraries, in which a type II restriction enzyme (e.g., Bsg I) is used to recircularize the plasmid after the concerted integration reaction, are underway. Because a few residues are simply replaced with an equivalent number, the mutagenesis should be less disruptive to overall protein topology and structure.

It is instructive to examine the insertions that were actually tolerated. Surprisingly, most (87%) of the ones present in the sorted and unsorted populations were present in the predicted transmembrane domains, again underscoring how little is known regarding the conformational requirements of CCR5. This is despite the fact that the insertion invariably leads to the addition of one or two charged residues (see Figure 2A). This was also true for the insertions present in all the cell populations, although there was a slight bias towards insertion in the carboxy terminal cytoplasmic domain (5 out of 50 compared to 21 out of 308). Note there were fully tolerated insertions in each of the four ECDs (Nt and ECL1–3), suggesting possible placement of additional epitope tags. This may be useful for structural studies using antibody reagents and fluorescence resonance energy transfer.

Co-receptor function of CCR5

Of nearly 150 insertions in CCR5 analyzed here, only a minority was informative regarding co-receptor function. Previous investigations had identified critical residues throughout CCR5, especially in the first ECD (Dragic, 1998; Farzan et al., 1998; Ross et al., 1998). Unfortunately, in that region (Nt) only a single insertion at K32 was informative here, and that was present in all populations. Thus, conclusions here are limited to examination of the informative mutations in the other domains. Of interest is the insertion in C184, which likely forms a covalent disulfide bond with C107 (Berger et al., 1999). C184 is recreated upon the oligonucleotide insertion, but the downstream residues are altered (see above). S186, which had been previously identified by substituting mouse residues into human CCR5 as being critical for co-receptor function (Ross et al., 1998), thus shifts position by five residues. In addition, other previously identified critical residues just C-terminal to S186, notably P189, Y190, and S191 (Kuhmann et al., 1997; Ross et al., 1998), would also be shifted by the same amount. It may thus be difficult to separate direct from indirect effects of the insertion at C184.

The same is true for the other three insertions (Y193, W196, and Q200) in ECL2 that specifically affected co-receptor function. The latter two resulted in protein truncations, which may have disrupted the loop structure or eliminated downstream determinants. For example, K203A enhances the D11A loss-of-function mutation for R5X4 envelopes (Doranz et al., 1997). No co-receptor function had previously been assigned to Y193. The effect observed at that residue may be direct or indirect, considering that it is a closed loop. There must be some flexibility in this loop structure, however, because insertions at R174 and E178 were well tolerated.

Can any conclusions be drawn from the truncating mutations at R229 and C230? The two major possibilities are that these truncating mutations are either structurally destabilizing (despite the fact they are expressed on the cell surface) or there are co-receptor determinants C-terminal, in the fourth ECD (ECL3). With regard to the latter, the CCR5 chimera studies have not been informative since the mouse sequence is identical to that of the human. Alanine-scanning mutagenesis has demonstrated the importance of D282 (in the context of D11A) for R5X4 strains (Doranz et al., 1997). This effect was not observed for the R5 strain JR-FL, although it is clear that the fourth ECD alone (in the context of murine CCR5) is insufficient for co-receptor function (Bieniasz et al., 1997). Of the mutants in the fourth ECD that we examined here, only one (4ECD-6) was informative in the sense that it was actually present on the cell surface. Minimal disruption of that loop (e.g., addition of one residue) led to undetectable surface expression. Based upon flow cytometry, 4ECD-6 was expressed on the cell surface, and although the percentage of positive cells was reduced 20-fold compared to wild-type, the amount of expression on

a per cell basis (as quantitated by MCF) was the same, if not higher. Thus, the fact that M-tropic viral titers were reduced approximately four log₁₀s is consistent with ECL3 playing a role in co-receptor function and probably deserves further exploration.

Previous investigations with mouse-human CCR5 chimeras revealed differing requirements for the extracellular loops for strains ADA, SF162, and BaL, with BaL being the most sensitive to the presence of murine sequences (Bieniasz et al., 1997; Picard et al., 1997). To a first approximation, we did not observe substantial differences between the three envelopes studied here, and whether that reflects the nature of the insertional mutagenesis versus the sensitivity of the PCR assay is unknown.

In summary, this report describes the application of genetic footprinting to the HIV co-receptor CCR5. Although the density of the insertion oligonucleotides was lower than expected, the important conclusions are that surface expression of CCR5 is relatively intolerant to 5 amino acid residue insertions, and critical co-receptor determinants lie within ECL2 and possibly ECL3. This is consistent with previous investigations using envelopes of R5 HIV isolates. It is entirely conceivable that a denser insertion library or alternatively a replacement genetic footprinting library may be more illuminating, depending upon the topological constraints of the gene under study.

Materials and methods

Plasmids

HIV-1 envelope expression plasmids pSM-ADA, pSM-BaL, and pSM-SF162 were gifts of Dan Littman (Skirball Institute, NYU Medical Center). MLV Gag-Pol expression construct pHIT60 (Soneoka et al., 1995) and pBabeCCR5-HA (Liu et al., 1996) were obtained from Alan Kingsman (Oxford) and Ned Landau (Salk Institute), respectively. HIV-blasti and HIV-Neo were constructed by substituting the coding region of alkaline phosphatase of the vector pHIV-APΔenvΔvifΔvpr (Sutton et al., 1998) with that of either the blasticidin resistance gene from pcDNA6/V5-HisA (Invitrogen) or SV-Neo (from pBabeNeo), respectively. HIV-IRES-eYFP has been reported previously (Schroers et al., 2000), as has the VSV G expression plasmid pME VSV G (Sutton et al., 1998).

Oligonucleotide insertion library of CCR5

Purified MuA transposase (gift of Harry Savilahti) was used to create a library of insertional mutations in pBabeCCR5-HA using modifications of described methods (Singh et al., 1997; Lauren et al., 2000). The donor oligonucleotide duplex used to create the insertions contained the recognition sequences for both the MuA transposase and MluI restriction endonuclease. This duplex was generated

by annealing together the oligonucleotides MuAMlu15A (5'-TGACGCGTCGCACGAAAAACGCGAAAGCGTTTC-ACGATAAATGCGAAAAC-3') and MuAMlu15B (5'-GTT-TTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCG-TGCGACGCGTCA-3'). The transposition reaction was performed by incubating 50 nM donor oligo duplex, 10 μ g pBabeCCR5-HA, and 100nM MuA transposase in a buffer containing 25 mM Tris-HCl pH 8.0, 144 mM NaCl, 10 mM MgCl₂, 100 μ g/ml bovine serum albumin, 15% (w/v) glycerol, 15% (v/v) dimethyl sulfoxide and 0.1% (w/v) Triton X-100 for 1 hour at 30°C. The reaction was extracted with phenol-chloroform and precipitated with ethanol. Plasmids that had undergone a concerted integration of two oligonucleotides (and therefore linear) were separated by horizontal agarose gel electrophoresis and purified. The five-base gap resulting from the transposition was repaired by nick translation with Taq DNA polymerase. The resulting molecules were cleaved with MluI, recircularized using T4 DNA ligase, and transformed into *E. coli* strain DH10B by electroporation. The resulting library contained over 450,000 independent transformants, each with the 15 bp insertion containing the MluI recognition sequence located at a random position in the pBabeCCR5-HA plasmid (see Figs. 1 and 2).

Cells and vector supernatants

293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS, from Gibco-BRL) and penicillin-streptomycin. Hos.T4 cells which express human CD4 under the control of an SV40 promoter were a gift of N. Landau and were maintained in complete DMEM supplemented with 40 μ g/ml mycophenolic acid (MPA, Sigma) with xanthine and hypoxanthine. Hos.T4.CCR5 cells that express both CD4 and wild-type CCR5 were a gift of D. Littman and were maintained in 5 μ g/ml puromycin (Sigma).

Pseudotyped HIV vector supernatants were produced by calcium-phosphate co-transfection of 293T cells (Sutton et al., 1998) with pHIV-blasti and the relevant envelope expression construct, titrated on Hos.T4.CCR5 cells using complete DMEM supplemented with 10 μ g/ml blasticidin S (Invitrogen). HIV-IRES-eYFP (VSV G) was prepared similarly, 1 ml of supernatant was added to targets (HOS.T4 cells expressing either wt or mutant CCR5) in 12-well plates, and 72 h later cells were subjected to flow cytometry, gating on the eYFP + cells. HIV-Neo (ADA) was similarly prepared, and 72 h after transduction, cells were positively selected by passaging at a ratio of 1:10 in complete DME supplemented with 1.0 mg/ml G418 (Invitrogen). Pseudotyped MLV vector supernatant was produced by calcium-phosphate co-transfection of 293T cells using pHIT60, pME VSV G, and the mutant pBabeCCR5-HA library or appropriate MLV vector construct. It was titrated using Hos.T4 cells as targets and also stored at -80°C.

Sorting for surface expression

Approximately 5×10^7 HOS.T4 cells were transduced with 2.4×10^6 IU of the pseudotyped MLV stock described above (m.o.i.=0.05), selected in DMEM containing both MPA and puromycin, and resistant colonies were pooled and expanded. Approximately 4×10^7 cells were incubated for 60 min at 4°C with 50 μ g of anti-HA mouse monoclonal antibody 12CA5 (Boehringer-Mannheim) in PBS containing 2% FCS (PBS-FCS). Cells were washed three times in PBS-FCS and bound to goat anti-mouse-coated magnetic beads (Miltenyi-Biotec). Cells and beads were then applied to a midi-MACS column (Miltenyi-Biotec). After washing extensively, bound cells were eluted according to the manufacturer's instructions.

Recovered cells were expanded in complete DMEM supplemented with MPA and puromycin. Flow cytometry was performed by staining approximately 10^6 cells using 2 μ g of 12CA5 and goat anti-mouse-fluorescein isothiocyanate (FITC) conjugate as a secondary antibody and analyzing them on a FACScan (Becton-Dickinson) using CellQuest software. Gates were established at the <1% level using cells stained with the secondary antibody alone.

Genetic footprinting

Positively sorted cells above were independently transduced with at least 10^6 IU of HIV-blasti(VSV G), HIV-blasti (ADA), HIV-blasti (BaL), and HIV-blasti (SF162). Transduced cells were selected in complete DMEM containing 10 μ g/ml blasticidin S, and expanded. Genomic DNA from approximately 2.5×10^7 cells was prepared by SDS-proteinase K lysis and phenol:chloroform extraction. Nested PCR was performed starting with 0.8 μ g of genomic DNA (approximately 100,000 cell equivalents) using the initial primer pair 5'-GACCCTCCCTTTATCCAGCCCT-CACTCCTT-3' (Gag-U) and 5'-CCGTGGGCTTGATC-TCGGTCATGGTAAGCT-3' (Puro-L). Conditions using 5 U of Taqpluslong (Boehringer-Mannheim) were 40 cycles of 94°C for 30 sec, 62°C for 30 sec, and 68°C for 90 sec to obtain a 1.4 kb product. Ten–20 ng of this product was used for the subsequent PCR reaction, using the following primer pairs (one of which was 5' end-labeled with biotin and the other 32-P): 5'-TTATACATCGGAGCCCTGCCAAA-3' (281U, 3' end at position 82) and 5'-AGCAAACACAG-CATGGACGACAG-3' (625L, 3' end at position 404), total length 367 bp; 5'-CCCTCCCTTTATCCAGCCCTCAC-3' (BgagU) and 5'-TGGCCAGGTTGAGCAGGTAGATG-3' (437L, 3' end at position 216), approximate total length 280 bp; 281U and 5'-ACGCAAACACAGCCACCACCC-AAGT-3' (693B, 3' end at position 472), total length 437 bp; 5'-GTCGTCCATGCTGTGTTTGCTTT-3' (627U, 3' end at position 428) and 5'-GCCCTGTGCCTCTTCTTCT-CATT-3' (915L, 3' end at position 694), total length 311; and 5'-GGTGTGCGAAATGAGAAGAAGAGG-3' (907U, 3' end at position 708) and 5'-GGGACTTTCACCCCTA-

ACTG-3' (1460L), approximate length 450 bp. Reaction conditions were 15 cycles of 94°C for 30 sec, 62°C for 30 sec, and 72°C for 30 sec using 5 U Taq DNA polymerase. Full-length PCR products were gel-isolated after size separation by horizontal agarose gel electrophoresis and then bound to streptavidin-agarose beads (SIGMA). After washing unbound material, beads were incubated with 10 U Mlu I for 60 min at 37°C. Beads were pelleted by centrifugation, and released DNA in the supernatant was precipitated with ethanol using glycogen as carrier. Mlu I-cleaved PCR products were electrophoresed on 40 cm 7M urea-6% polyacrylamide 0.4mm sequencing gels, along with the homologous dideoxy sequence. For quantitation of specific products, Cerenkov counts were obtained for the full length PCR products prior to binding to streptavidin-agarose beads (for normalization purposes), and individual bands were quantitated using a Betagen blot analyzer system.

Construction of mutations in the fourth ECD of CCR5 and analysis of surface expression

4ECD-1 was constructed by PCR using pBabeCCR5-HA as a template with the following primers 5'-GGG-AATTTCGGGCATATGCAGGTGACAGAGACTCTTGG-3' (Eco 4U) and 5'-GGGTCGACTCACAAGCCCACAGATATTCCTG-3' (Eco 4L). The ~300 bp product was first cloned into pCR 2.1-TOPO (Invitrogen), sequence confirmed, and then directionally ligated into pBabeCCR5-HA pre-cleaved with Eco RI and Sal I. 4ECD-2 was constructed by cleaving 4ECD-1 with Eco RI and Nde I and inserting a duplex oligonucleotide, with top strand sequence 5'-AATTCAATAATTGCAGTAGCTCTAACAGGTTGGACCAAGCTCA-3'. 4ECD-3 was constructed by cleaving 4ECD-1 with Eco RI and Nde I and inserting a duplex oligonucleotide with top strand sequence of 5'-AATTCTTTGGCC-TGAATAATTGCAGTAGCTCTAACAGGTTGGACCAAGCTCA-3'. 4ECD-4 was constructed by PCR using pBabeCCR5-HA as a template and the following primers: 5'-GGGAATTCAATAATTGCAGTAGCTCTAACAGGTTGG-3' and the above-mentioned Eco 4L primer. The ~285 bp product was cloned into pBabeCCR5-HA as described for 4ECD-1. 4ECD-5 and 4ECD-6 were constructed by cleaving pBabeCCR5-HA with Eco RI and non-directionally ligating a duplex oligonucleotide with top strand sequence of 5'-AATTGTCGACTG-3'. Both forward and reverse orientation clones were confirmed by DNA sequencing. The amino acid sequences of the mutant ECDs are shown in Table 1.

To examine cell surface expression of the 4ECD mutants, HOS.CD4 cells were transduced with either pBabeCCR5-HA or pBabeCCR5-HA-4ECD-6, both pseudotyped with VSV G. Stable transductants were selected using DME complete medium supplemented with 5 µg/ml puromycin. For surface staining, washed cells were incubated with PE-conjugated anti-human CCR5 mouse monoclonal antibody (FAB182P, R&D Systems) according to the manufacturer's recommendations and analyzed on a Becton-Dickinson FACScan. Gating was

established so that only 0.1% of the HOS.CD4 parental cells were positive.

Acknowledgments

We would like to thank Dr. Patrick Brown (Stanford) for his support of this project in its early stages, and Drs. Ned Landau (Salk), Dan Littman (NYU), Sue and Alan Kingsman (Oxford), and Harry Savilahti (NIH) for generous reagent gifts. This work was funded in part by the Baylor Center for AIDS Research Core Support Grant number AI36211 from NIAID, and R.Q. was supported by T32 AI 07456. R.E.S. is an Edward Mallinckrodt, Jr. Foundation Scholar.

References

- Alkhatib, G., Locati, M., Kennedy, P.E., Murphy, P.M., Berger, E.A., 1997. HIV-1 coreceptor activity of CCR5 and its inhibition by chemokines: independence from G protein signaling and importance of coreceptor downmodulation. *Virology* 234, 340–348.
- Atchison, R.E., Gosling, J., Monteclaro, F.S., Franci, C., Digilio, L., Charo, I.F., Goldsmith, M.A., 1996. Multiple extracellular elements of CCR5 and HIV-1 entry: dissociation from response to chemokines. *Science* 274, 1924–1926.
- Berger, E.A., Murphy, P.M., Farber, J.M., 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* 17, 657–700.
- Berson, J.F., Doms, R.W., 1999. Structure-function studies of the HIV-1 coreceptors. *Semin. Immunol.* 10, 237–248.
- Bieniasz, P.D., Fridell, R.A., Aramori, I., Ferguson, S.S., Caron, M.G., Cullen, B.R., 1997. HIV-1-induced cell fusion is mediated by multiple regions within both the viral envelope and the CCR-5 co-receptor. *Embo. J.* 16, 2599–2609.
- Brown, P.O., Bowerman, B., Varmus, H.E., Bishop, J.M., 1987. Correct integration of retroviral DNA in vitro. *Cell* 49, 347–356.
- Doranz, B.J., Lu, Z.H., Rucker, J., Zhang, T.Y., Sharron, M., Cen, Y.H., Wang, Z.X., Guo, H.H., Du, J.G., Accavitti, M.A., Doms, R.W., Peiper, S.C., 1997. Two distinct CCR5 domains can mediate coreceptor usage by human immunodeficiency virus type 1. *J. Virol.* 71, 6305–6314.
- Dragic, T., 2001. An overview of the determinants of CCR5 and CXCR4 co-receptor function. *J. Gen. Virol.* 82, 1807–1814.
- Dragic, T., Trkola, A., Lin, S.W., Nagashima, K.A., Kajumo, F., Zhao, L., Olson, W.C., Wu, L., Mackay, C.R., Allaway, G.P., Sakmar, T.P., Moore, J.P., Maddon, P.J., 1998. Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J. Virol.* 72, 279–285.
- Farzan, M., Choe, H., Vaca, L., Martin, K., Sun, Y., Desjardins, E., Ruffing, N., Wu, L., Wyatt, R., Gerard, N., Gerard, C., Sodroski, J., 1998. A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J. Virol.* 72, 1160–1164.
- Kuhmann, S.E., Platt, E.J., Kozak, S.L., Kabat, D., 1997. Polymorphisms in the CCR5 genes of African green monkeys and mice implicate specific amino acids in infections by simian and human immunodeficiency viruses. *J. Virol.* 71, 8642–8656.
- Laurent, L.C., Olsen, M.N., Crowley, R.A., Savilahti, H., Brown, P.O., 2000. Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. *J. Virol.* 74, 2760–2769.
- Lee, B., Sharron, M., Blanpain, C., Doranz, B.J., Vakili, J., Setoh, P., Berg, E., Liu, G., Guy, H.R., Durell, S.R., Parmentier, M., Chang, C.N.,

- Price, K., Tsang, M., Doms, R.W., 1999. Epitope mapping of CCR5 reveals multiple conformational states and distinct but overlapping structures involved in chemokine and coreceptor function. *J. Biol. Chem.* 274, 9617–9626.
- Littman, D.R., 1998. Chemokine receptors: keys to AIDS pathogenesis? *Cell* 93, 677–680.
- Liu, R., Paxton, W.A., Choe, S., Ceradini, D., Martin, S.R., Horuk, R., MacDonald, M.E., Stuhlmann, H., Koup, R.A., Landau, N.R., 1996. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 86, 367–377.
- Olson, W.C., Rabut, G.E., Nagashima, K.A., Tran, D.N., Anselma, D.J., Monard, S.P., Segal, J.P., Thompson, D.A., Kajumo, F., Guo, Y., Moore, J.P., Maddon, P.J., Dragic, T., 1999. Differential inhibition of human immunodeficiency virus type 1 fusion, gp120 binding, and CC-chemokine activity by monoclonal antibodies to CCR5. *J. Virol.* 73, 4145–4155.
- Picard, L., Simmons, G., Power, C.A., Meyer, A., Weiss, R.A., Clapham, P.R., 1997. Multiple extracellular domains of CCR-5 contribute to human immunodeficiency virus type 1 entry and fusion. *J. Virol.* 71, 5003–5011.
- Pryciak, P.M., Varmus, H.E., 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69, 769–780.
- Rabut, G.E., Konner, J.A., Kajumo, F., Moore, J.P., Dragic, T., 1998. Alanine substitutions of polar and nonpolar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage- and dualtropic isolates of human immunodeficiency virus type 1. *J. Virol.* 72, 3464–3468.
- Ross, T.M., Bieniasz, P.D., Cullen, B.R., 1998. Multiple residues contribute to the inability of murine CCR-5 to function as a coreceptor for macrophage-tropic human immunodeficiency virus type 1 isolates. *J. Virol.* 72, 1918–1924.
- Rossi, D., Zlotnik, A., 2000. The biology of chemokines and their receptors. *Annu. Rev. Immunol.* 18, 217–242.
- Rothenberg, S.M., Olsen, M.N., Laurent, L.C., Crowley, R.A., Brown, P.O., 2001. Comprehensive mutational analysis of the Moloney murine leukemia virus envelope protein. *J. Virol.* 75, 11851–11862.
- Rucker, J., Samson, M., Doranz, B.J., Libert, F., Berson, J.F., Yi, Y., Smyth, R.J., Collman, R.G., Broder, C.C., Vassart, G., Doms, R.W., Parmentier, M., 1996. Regions in beta-chemokine receptors CCR5 and CCR2b that determine HIV-1 cofactor specificity. *Cell* 87, 437–446.
- Sallusto, F., Mackay, C.R., Lanzavecchia, A., 2000. The role of chemokine receptors in primary, effector, and memory immune responses. *Annu. Rev. Immunol.* 18, 593–620.
- Samson, M., 1996. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* 382, 722–725.
- Schroers, R., Sinha, I., Segall, H., Schmidt-Wolf, L.G.H., Rooney, C.M., Brenner, M.K., Sutton, R.E., Chen, S.-Y., 2000. Transduction of human PBMC-derived dendritic cells and macrophages by an HIV-1-based lentiviral vector system. *Molecular Therapy* 1, 171–179.
- Singh, I.R., Crowley, R.A., Brown, P.O., 1997. High-resolution functional mapping of a cloned gene by genetic footprinting. *Proc. Natl. Acad. Sci. USA* 94, 1304–1309.
- Soneoka, Y., Cannon, P.M., Ramsdale, E.E., Griffiths, J.C., Romano, G., Kingsman, S.M., Kingsman, A.J., 1995. A transient three-plasmid expression system for the production of high titer retroviral vectors. *Nucleic Acids Res.* 23, 628–633.
- Sutton, R.E., Wu, H.T., Rigg, R., Bohnlein, E., Brown, P.O., 1998. Human immunodeficiency virus type 1 vectors efficiently transduce human hematopoietic stem cells. *J. Virol.* 72, 5781–5788.
- Wu, L., LaRosa, G., Kassam, N., Gordon, C.J., Heath, H., Ruffing, N., Chen, H., Humblas, J., Samson, M., Parmentier, M., Moore, J.P., Mackay, C.R., 1997. Interaction of chemokine receptor CCR5 with its ligands: multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding. *J. Exp. Med.* 186, 1373–1381.